Identification of Non–Small Cell Lung Cancer Sensitive to Systemic Cancer Therapies Using Radiomics

Laurent Dercle1,2, Matthew Fronheiser3, Lin Lu1, Shuyan Du3, Wendy Hayes3, David K. Leung3, Amit Roy4, Julia Wilkerson5, Pingzhen Guo1, Antonio T. Fojo6, Lawrence H. Schwartz1, and Binsheng Zhao1

Purpose: Using standard-of-care CT images obtained from patients with a diagnosis of non–small cell lung cancer (NSCLC), we defined radiomics signatures predicting the sensitivity of tumors to nivolumab, docetaxel, and gefitinib.

Experimental Design: Data were collected prospectively and analyzed retrospectively across multicenter clinical trials [nivolumab, n = 92, CheckMate017 (NCT01642004), CheckMate063 (NCT01721759); docetaxel, n = 50, CheckMate017; gefitinib, n = 46, (NCT00588445)]. Patients were randomized to training or validation cohorts using either a 4:1 ratio (nivolumab: 72T:20V) or a 2:1 ratio (docetaxel: 32T:18V; gefitinib: 31T:15V) to ensure an adequate sample size in the validation set. Radiomics signatures were derived from quantitative analysis of early tumor changes from baseline to first on-treatment assessment. For each patient, 1,160 radiomics features were extracted from the largest measurable lung lesion. Tumors were classified as treatment sensitive or insensitive; reference standard was median progression-free survival (NCT01642004, NCT01721759) or surgery (NCT00588445). Machine learning was implemented to select up to four features to develop a radiomics signature in the training datasets and applied to each patient in the validation datasets to classify treatment sensitivity.

Results: The radiomics signatures predicted treatment sensitivity in the validation dataset of each study group with AUC (95 confidence interval): nivolumab, 0.77 (0.55–1.00); docetaxel, 0.67 (0.37–0.96); and gefitinib, 0.82 (0.53–0.97). Using serial radiographic measurements, the magnitude of exponential increase in signature features deciphering tumor volume, invasion of tumor boundaries, or tumor spatial heterogeneity was associated with shorter overall survival.

Conclusions: Radiomics signatures predicted tumor sensitivity to treatment in patients with NSCLC, offering an approach that could enhance clinical decision-making to continue systemic therapies and forecast overall survival.

Introduction

Selecting patients for targeted therapies or immunotherapy is crucial to match individuals to the treatment most likely to benefit them. In patients with a diagnosis of non–small cell lung cancer (NSCLC), personalization of therapy currently relies on pretreatment biomarkers acquired in a tumor biopsy taken at baseline. Tumor biopsies are used to perform genomic analyses to identify patterns present in images. These quantitative imaging biomarkers, such as EGFR and anaplastic lymphoma kinase, ALK, as well as the expression of proteins that might help predict sensitivity to immunotherapy (e.g., programmed cell death 1 ligand, PD-L1). These measures are typically limited to a single biopsy sample, are difficult to perform repeatedly, and thus cannot capture the spatial and temporal heterogeneity of disease.

Progress in artificial intelligence (AI) has transformed the field of radiology, especially radiomics. Radiomics depends on the quantitative transformation of images into comprehensive datasets that enables high-throughput data mining and automated analysis of patterns present in images. These quantitative imaging biomarkers, defined a priori using mathematical formulas, could guide treatment decision. Radiomics features are calculated by algorithmic analysis of tumor images and have been linked to characteristics of NSCLC. AI can be trained to recognize clinically relevant patterns on CT images and perform a “digital biopsy” of the imaging phenotype of the entire tumor volume. Quantitative imaging features have been associated with therapeutically actionable mutations, such as EGFR mutation status (1–15). Early volumetric assessment of variation in imaging phenotype on CT scan (16–23) has been shown to predict the biologic activity of targeted therapies such as anti-EGFR agents (17, 23). Finally, the imaging phenotype of NSCLC has been linked to patients’ outcome (23, 24–30) by predicting metastasis risk, recurrence risk, gross residual disease, and survival.

Currently, the main strategy for response assessment is based on a radiologist’s evaluation of the changes in the size and the appearance of new tumor lesions. In the case of “targeted therapies,” shrinkage of target lesions is considered a hallmark of dependency on the targeted pathway and consequently of treatment sensitivity. With low variation due to lesion sampling, acquisition protocol, and observer effect, tumor shrinkage was a robust tool in the chemotherapy era to assess anticancer treatment efficacy, and is widely employed using RECIST 1.1 (31–33). In immune oncology, the utility of RECIST is more limited due to atypical patterns of response and progression on medical imaging. Despite the use of new imaging response assessment criteria, an unmet clinical need for improved assessment remains, which justifies the need for alternative approaches.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Laurent Dercle, Columbia University Medical Center, 168th Street, New York, NY 10032. Phone: 9172830908; E-mail: laurent.dercle@gmail.com

Clin Cancer Res 2020;26:2151–62

doi: 10.1158/1078-0432.CCR-19-2942

©2020 American Association for Cancer Research.
Translational Relevance

New patterns of response and progression have been observed in patients treated with immunotherapy, such as pseudoprogression and hyperprogression, prompting the need for alternative metrics for response assessment and therapeutic decision-making. Radiomic signatures, derived from quantitative, artificial intelligence-based analysis of standard-of-care CT images, offer the potential to enhance clinical decision-making as on-treatment markers of efficacy. In patients with non–small cell lung cancer treated with a wide spectrum of systemic cancer therapies (nivolumab, docetaxel, or gefitinib), radiomic signatures detected early changes from baseline to first on-treatment tumor assessment that were associated with sensitivity to treatment. Using serial radiographic measurements, we observed that an exponential increase in signature features deciphering tumor volume, invasion of tumor boundaries, or tumor spatial heterogeneity was associated with treatment insensitivity and shorter overall survival. This indicates radiomic signatures offer an approach that could guide clinical decision-making to continue or modify systemic therapies.

The investigation of pretreatment biomarkers to identify patients who might benefit from immunotherapy has gained traction within the research community. The unique challenge is unconventional patterns of response and progression including hyperprogression and pseudoprogression. As ancillary studies, given the clear implication of onbiologic progression, it was our hypothesis that extending our imaging evaluation to include pretreatment imaging features as well as serial radiographic measurements might be an untapped source of complementary prognostic information.

Our focus was on advanced/metastatic NSCLC because it is a unique candidate for implementing a radiomics approach. First, there is a strong clinical need because lung cancer is the second most common cancer and a leading cause of cancer death for men and women. Second, the segmentation of lung lesion can be easily implemented in clinical routine because the healthy air-filled lung parenchyma is the most hypodense human tissue. Third, from a biological perspective, the paradigm of response was developed in cytotoxic chemotherapies, which generates a need to explore change in imaging phenotype induced by other types of systemic treatments such as targeted therapy and immunotherapy. Finally, radiographic response evaluation is a standard of care, and the response rate is known to be sufficient to make the results meaningful (34).

We aimed to explore whether AI techniques may be of clinical utility to oncologists as on-treatment markers of efficacy to help decide which patients should continue treatment. To this end, we evaluated the performance of treatment-specific radiomics signatures measured at baseline and at the first response assessment in patients with NSCLC receiving treatment with either of three cancer therapies: an immunotherapy blocking a negative regulator of T-cell activation and response (nivolumab, a monoclonal IgG4 antibody targeting PD-1), a chemotherapy targeting microtubules (docetaxel), and a molecular targeted therapy interrupting EGFR signaling (gefitinib).

Materials and Methods

Our primary objective was to train and validate three on-treatment signatures to detect NSCLC tumors sensitive to nivolumab, docetaxel, or gefitinib using a quantitative analysis of early tumor changes from baseline to first on-treatment tumor assessment on serial CT scans. Our secondary objective was to test whether features were generalizable across treatment arms.

Selection of eligible trials

The following inclusion criteria were used for trial eligibility: completed study; primary tumor type with a large proportion of measurable, quantifiable disease (NSCLC); ≥40 patients accrued; FDA-approved drug; and CT images centrally collected and archived. We selected completed trials evaluating three different types of drug classes (an immunotherapy, a cytotoxic chemotheraphy, and a molecularly targeted agent) so that imaging metrics could be studied across a range of different therapies.

Participants

We retrospectively analyzed three NSCLC clinical trials using three treatments (nivolumab, docetaxel, and gefitinib) and 188 patients. Patients treated with nivolumab (n = 92) had enrolled in the NCT01642004 and NCT01721759 multicenter phase II–III trials. Patients who received docetaxel (n = 50) had been treated in the NCT01642004 trial. Patients prescribed gefitinib (n = 46) received the agent on NCT00588445, a single arm phase II trial in NSCLC that sought to correlate the radiographic response induced by gefitinib with mutations in the protein-tyrosine kinase domain of the EGFR gene. Patient characteristics are summarized in Table 1 and in Supplementary SI.1 and SI.2.

The investigators of these clinical trials obtained written informed consent from the patients. The studies were conducted in accordance with recognized ethical guidelines (Declaration of Helsinki), and the studies were approved by an Institutional Review Board.

Data were collected up to the completion date of the clinical trials. Patients were randomly assigned to either training (T) or validation (V) sets using either a 4:1 ratio (nivolumab: 72T:20V) or a 2:1 ratio (docetaxel: 32T:18V; gefitinib: 31T:15V) to ensure an adequate sample size in the validation set. We estimated that a sample size of a minimum of 14 patients was required in the validation set based on the following input and assumption: a type I error of 0.05, a power of 0.8, an AUC of 0.85, and an allocation ratio of 1 (35).

In NCT01642004 and NCT01721759, CT scan imaging was performed at baseline and again every 8 weeks until disease progression or withdrawal. In NCT00588445, CT scan imaging occurred at baseline and at 3 weeks, just prior to day-23 surgery. Patients with missing data were excluded. Additional trial details are included in Supplementary SI.1.

Quality of CT scan acquisition

We selected 188 patients out of 264 eligible patients (Fig. 1, Table 1) based on eligibility criteria ensuring improved imaging quality for this quantitative retrospective analysis: (i) measurable lung lesions according to RECIST 1.1 at baseline; (ii) no significant imaging artifacts; (iii) lung reconstruction kernel; (iv) pixel spacing <1 mm; (v) slice thickness <10 mm; and (vi) CT scans available at baseline and first response evaluation (landmark).

The radiomics quality score (RQS) has been proposed as a guideline to evaluate the quality of radiomics studies (36). For the current study, the RQS was estimated to be 28 out of 36 points (78%). More information about CT scan characteristics, CT scan quality (37–39), and RQS can be found in Supplementary SI.1. A depiction of the radiomics workflow is shown in Fig. 1.
Lesion segmentation and feature extraction

The largest measurable lung tumor present at baseline was segmented in the baseline and in the first on-treatment response assessment CT scan (nivolumab, 8 weeks; docetaxel, 8 weeks; gefitinib, 3 weeks) in all patients that met the inclusion criteria. Segmentation was performed using an algorithm developed in-house that enables semiautomatic creation of contours on all available CT scans (40). Imaging features were extracted from lung tumors using a priori definitions of radiomics features (38). In total, 1,160 quantitative image features were extracted from the images of each lesion from both the baseline and the first on-treatment CT scan. Delta radiomics features were calculated to characterize the early changes in the features. Full details of lesion segmentation and feature extraction can be found in Supplementary SI.2 and SI.3.

Signature building in each treatment arm

In the training set of each cohort, we developed a multivariable prediction model, i.e., the signature, to predict treatment sensitivity. Using machine learning, quantitative imaging features were combined by high-throughput mining to build the signature. In keeping with our ultimate aim to generate simple to use, noninvasive clinical decision tools using standard of care CT scan images, the radiomics signature ranging from 0 (highest treatment sensitivity) to 1 (highest treatment insensitivity) for each patient was based on the analysis of the change of the largest measurable lung lesion identified at baseline on CT scan.

In the implementation, a "coarse" to "fine" strategy was developed to select optimal features from the large number of the extracted quantitative image features to build the signature. The coarse selection approach consists of reproducibility analysis, redundancy analysis, and feature ranking that eliminate those nonreproducible, redundant,
and noninformative features. The fine selection approach was composed of "forward" search and feature combination, aiming to select the most significant features to build the best predictive model. To prevent overfitting, up to the top four features (in terms of prediction importance outputted by the machine-learning algorithm; ref. 41) of the identified image features in the best predictive model developed in the training set were integrated in the signature. Full details of the model building can be found in Supplementary SI.4.

Due to the limited number of patients available in the gefitinib cohort (n = 46 patients), we defined an alternative strategy for signature building. We trained (n = 202 patients) and validated (n = 100 patients) a four-feature treatment sensitivity signature (Random Forest algorithm; ref. 41) in a cohort of patients with metastatic colorectal cancer treated with anti-EGFR monoclonal antibodies. To prevent overfitting, the four features identified in the colorectal cancer signature were transferred to be calibrated to predict treatment sensitivity in the training set of the gefitinib cohort.

The radiomics signature was trained to predict the tumor sensitivity to systemic anticancer treatment. All patients with cancer were divided into two groups: sensitive and insensitive to treatment. In patients with NSCLC treated with nivolumab and docetaxel, the reference standard to determine treatment sensitivity was median progression-free survival.
survival (PFS). The reference standard for gefitinib-treated patients was derived from the analysis of the surgical specimen 2 days after stopping 21 days of gefitinib therapy. The independent validation dataset, consisting of unseen data that were not used for training, was used to evaluate the performance of the signature. Further details for the reference standards are provided in Supplementary SI5.1.

Generalizability of the features
Our primary objective was to evaluate one single lesion at two timepoints (baseline and first on-treatment assessment). As an ancillary study, we evaluated the generalizability of imaging features and compared them with alternative outcome measures. To this end, we used all measurable lesions when we studied the clinical value of imaging features using baseline measurement (one timepoint) or serial radiographic measurement (≥two timepoints).

Estimating rates of tumors with exponential changes in radiomics features using serial radiographic measurements
We attempted to reduce the risk of type I error/overfitting and to generalize the features identified in the three radiomics signatures (nivolumab, docetaxel, and gefitinib). To this end, we modeled the evolution of these features across time using serial radiographic measurements of tumors. We evaluated if these models could be used to understand disease behavior during treatment, compare study interventions, and forecast overall survival (OS; ref. 42).

Previous studies have demonstrated the simultaneous occurrence of two processes in the overwhelming majority of tumors: exponential growth of the treatment-insensitive fraction of the tumor at a rate described by a growth rate constant designated g for growth, and exponential regression of the treatment-sensitive portion of the tumor at a rate described by a regression rate constant designated d for decay. Both processes occur exponentially, and the rates of growth and regression can be estimated using simple mathematical equations. Using the same equations, we were able to estimate the rates of exponential change in the radiomics features over time and based on the rate of change for each radiomic feature assigned tumors into one of three categories: (i) those in which the designated feature was observed to only exponentially grow/increase or become more prominent during treatment (gx); (ii) those in which the analyzed feature only decreased or disappeared exponentially in quantity during therapy (dx); and (iii) those in which the data were best described by an equation that considered that there had occurred concurrent exponential increase and disappearance of the radiomic feature examined, as portions of the tumor sensitive to the therapy disappeared while those resistant to the treatment increased in abundance (gd; ref. 42). Rates of growth or decay of each feature were obtained by using all available radiographic measurements from baseline to a landmark analysis at 8 months. The differences in OS between patients with g above median and below median were analyzed using a Cox proportional hazards model and log-rank test (Kaplan–Meier analysis) in which landmark analysis divided the follow-up time at 8-month time point and a P value less than 0.05 was considered significant.

Baseline imaging predictors of radiographic progression
We evaluated whether baseline radiomics features could predict the best overall response of individual lung lesions. To this end, we classified all measurable lung lesions in the nivolumab cohort into two categories: progressive versus nonprogressive per iRECIST criteria (43). Supplementary SI6.3 contains additional methodology details.

Statistical analysis
Statistical analysis was conducted using Matlab2016a and SPSS23.0. The reported P values were two-sided, with the level for statistical significance set at α = 0.05. The performance of the model was evaluated by computing the area under the ROC curve.

Results

Performance of radiomics signature: nivolumab
In the training set (n = 72), the tumors of 28 patients were classified as sensitive to nivolumab. The nivolumab radiomics signature included four delta-radiomics features characterizing the change in tumor volume, heterogeneity, and margin sharpness: delta-Volume, delta-GLCM IMC1 (Gray-Level Co-Occurrence Matrix), delta-DWT1 (Discrete Wavelet Transform), and delta-sigmoid slope (Table 1). The nivolumab signature achieved an AUC (95 confidence interval (CI)) of 0.80 (0.69–0.80; P < 10⁻⁵) in its ability to identify the sensitivity of a patient’s tumor to nivolumab.

In the validation set (n = 20), the tumors of 5 patients were classified as sensitive to nivolumab. The performance of the nivolumab signature in the validation set was AUC (95 CI) of 0.77 (0.55–1.00). Using Kaplan–Meier plots, the estimated median PFS of the overall population (95 CI) was 2.1 (1.3–2.9) months (Fig. 2A). The estimated median PFS (95 CI) was 2.0 (1.8–2.2) versus 6.3 (4.0–8.6) months for patients with (n = 7 patients) or without (n = 35 patients) a high-risk nivolumab signature (signature > 0.5) respectively (P < 10⁻⁴).

Performance of radiomics signature: docetaxel
In the training set (n = 32), the tumors of 5 patients were classified as sensitive to docetaxel. The radiomics signature for the docetaxel cohort was a single delta-radiomics feature, delta-Volume (Table 1). The performance of the docetaxel signature was AUC (95 CI) of 0.68 (0.38–0.98).

In the validation set (n = 18), the tumors of 6 patients were classified as sensitive to docetaxel, and the performance of the docetaxel radiomics signature was AUC (95 CI) of 0.67 (0.37–0.96). Using Kaplan–Meier analyses, we observed that compared with tumors with a high-risk docetaxel signature (signature > 0.5, n = 39 patients) those without the high-risk signature (signature <= 0.5, n = 35 patients) had a longer estimated median PFS (95 CI) of 6.2 (5.5–7.0) months as compared with 2.1 (2.0–2.3) months (P < 0.001; Fig. 2B).

Performance of radiomics signature: gefitinib
In the training set (n = 31), the tumors of 13 patients were classified as sensitive to gefitinib. The gefitinib signature was composed of four delta-radiomics features characterizing the change in tumor shape and heterogeneity: delta-Shape-SI4, delta-LOG-X-Entropy, delta-LOG-Z-Entropy, and delta-GTDM-Contrast (Table 1). The performance of the gefitinib signature in the training set was AUC (95 CI) of 0.81 (0.61–0.92).

In the validation set (n = 15), the tumors of 7 patients were classified as sensitive to gefitinib. The performance of the gefitinib radiomics signature was AUC (95 CI) of 0.82 (0.53–0.97). Table 1 includes a summary of the patient information, treatment-specific radiomics signatures, and performance metrics.

Estimating rates of tumors with exponential changes in radiomics features using serial radiographic measurements
In addition to assessing the radiomic features at baseline and in the first on-treatment tumor assessment, we were also interested in examining the kinetics of change in these same radiomic features as
treatment was administered. To do this, we used the values of the eight radiomic features mentioned above from serial scans to assess which equation previously shown to describe the rates of change in tumor volume over time could best describe the kinetics of change in the radiomic features with therapy. Table 2 shows that in the majority of the 224 patients with data available for analysis, simple mathematical equations could be used to describe three categories for each radiomics feature in the analyzed tumors: those with only exponential increase in the studied radiomics feature (\( g \)), those with only exponential decrease in the analyzed feature (\( d \)), and those in which the kinetic change in the studied radiomics feature was best described by an equation that included concurrent exponential increase and reduction (\( g \) and \( d \)). Figure 3 shows the distribution (i.e., bimodal, trimodal) of \( g \) and \( d \) values for each radiomics feature.

To simplify the analysis and its interpretation, the eight features mentioned above were divided into three categories: tumor burden (volume), tumor heterogeneity (LOG-X-Entropy, LOG-Z-Entropy, GTDM-Contrast, GLCM-IMC1, GTDM-contrast), and boundaries (Shape-SI4, sigmoid-slope). Tumor burden was established with either unidimensional or bidimensional measurements. The average per-tumor volume over time could best describe the kinetics of change in the designated radiomic feature (\( g \)) of the radiomic features included in the signatures was associated with OS in patients from both treatment groups (C–E, pooled groups). The magnitude of exponential increase in tumor volume (\( g_{\text{Volume}} \)), tumor spatial heterogeneity (\( g_{\text{GLCM-IMC1}} \) or \( g_{\text{Shape-SI4}} \)), or boundary irregularity (\( g_{\text{Shape-SI4}} \)) was associated with shorter OS.

Figure 2 shows the distribution of \( g \) and \( d \) values for each radiomics feature. To describe the rates of change in tumor volume over time, we observed that the rate of exponential increase (\( g \)) was associated with OS in patients from both treatment groups (Figure 2, Dercle et al.).

Probability of PFS and OS over time as a function of signature score and signature features. Prolonged PFS was observed in patients with low-risk/treatment-sensitive signatures (\( <0.5 \)) in both treatment study groups (Fig. 2) using baseline and 8-week CT scans. In the nivolumab cohort (A), median PFS (95% CI) was 2.0 months (1.8–2.2) for patients whose tumors had a signature score \( >0.5 \) (predicted insensitivity, \( n = 57 \)) and 6.3 months (4.0–8.6) for patients with a signature score \( \leq 0.5 \) (predicted sensitivity, \( n = 35 \); \( P < 10^{-5} \)). In the docetaxel cohort (B), median PFS (95% CI) was 2.1 months (2.0–3.3) for patients whose tumors had a signature score \( >0.5 \) (predicted insensitivity, \( n = 39 \)) and 6.2 months (5.5–7.0) for signature score \( \leq 0.5 \) (predicted sensitivity, \( n = 11 \); \( P < 10^{-5} \)). Using serial radiographic measurements and a landmark at 8-month after drug initiation, we observed that the rate of exponential increase (\( g \)) of the radiomic features included in the signatures was associated with OS in patients from both treatment groups (C–E, pooled groups). The magnitude of exponential increase in tumor volume (\( g_{\text{Volume}} \)), tumor spatial heterogeneity (\( g_{\text{GLCM-IMC1}} \) or \( g_{\text{Shape-SI4}} \)), or boundary irregularity (\( g_{\text{Shape-SI4}} \)) was associated with shorter OS.

Baseline imaging predictors of radiographic progression

On a per-lesion analysis, the best overall response was objective radiographical progression in 136 (36.4%) lesions treated with nivolumab and evaluated in at least two timepoints (baseline and 8 weeks). The best overall response in other lesions (63.6%) was pseudoprogression (\( n = 4 \; 1.1 \%), response (\( n = 26 \); 7.0%), and stability (\( n = 207 \); 55.5%). An analysis of baseline radiomics features of these lung lesions demonstrated that the best baseline predictors of progression were features associated with tumor heterogeneity. The imaging feature Shape-SI4 provided clinically useful information across four treatment arms. Shape-SI4, originally identified in patients with colorectal cancer (44), was transferred to predict sensitivity to gefitinib in patients with NSCLC. Shape-SI4 was associated with OS in NSCLC treated with docetaxel and nivolumab. See additional information in Supplementary SI.6 and SI.4.
Table 2. Summary of the percentage of patients fitting $g$ and $d$ models for the eight radiomics features discovered in the three signatures.

<table>
<thead>
<tr>
<th></th>
<th>Tumor burden</th>
<th>Heterogeneity</th>
<th>Boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average$^a$</td>
<td>Uni</td>
<td>Bi</td>
</tr>
<tr>
<td><strong>Nivolumab</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients analyzed, n (%)</td>
<td>129 (81.6)</td>
<td>122 (77.2)</td>
<td>128 (81.0)</td>
</tr>
<tr>
<td></td>
<td>10 (6.6)</td>
<td>11 (7)</td>
<td>9 (5.7)</td>
</tr>
<tr>
<td></td>
<td>77 (48.5)</td>
<td>68 (43)</td>
<td>75 (47.5)</td>
</tr>
<tr>
<td></td>
<td>37 (23.6)</td>
<td>40 (25.3)</td>
<td>39 (24.7)</td>
</tr>
<tr>
<td><strong>Docetaxel</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients analyzed, n (%)</td>
<td>49 (74.7)</td>
<td>46 (69.7)</td>
<td>53 (80.3)</td>
</tr>
<tr>
<td></td>
<td>3 (4.0)</td>
<td>4 (6.1)</td>
<td>2 (3)</td>
</tr>
<tr>
<td></td>
<td>28 (42.9)</td>
<td>23 (34.8)</td>
<td>32 (48.5)</td>
</tr>
<tr>
<td></td>
<td>17 (25.8)</td>
<td>18 (27.3)</td>
<td>18 (27.3)</td>
</tr>
<tr>
<td><strong>Geftinib</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients analyzed, n (%)</td>
<td>17 (25.3)</td>
<td>20 (30.3)</td>
<td>13 (19.7)</td>
</tr>
<tr>
<td></td>
<td>12 (18.7)</td>
<td>19 (28.8)</td>
<td>8 (12.1)</td>
</tr>
<tr>
<td></td>
<td>4 (6.6)</td>
<td>1 (1.5)</td>
<td>5 (7.6)</td>
</tr>
<tr>
<td><strong>Gefitinib</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients analyzed, n (%)</td>
<td>5 (10.9)</td>
<td>4 (8.7)</td>
<td>5 (10.9)</td>
</tr>
<tr>
<td></td>
<td>0.7 (1.5)</td>
<td>1 (2.2)</td>
<td>1 (2.2)</td>
</tr>
<tr>
<td></td>
<td>0.3 (0.7)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>4 (8.7)</td>
<td>3 (6.5)</td>
<td>4 (8.7)</td>
</tr>
<tr>
<td>Patients not analyzed, n (%)</td>
<td>41 (89.1)</td>
<td>42 (91.3)</td>
<td>41 (89.1)</td>
</tr>
<tr>
<td></td>
<td>0.7 (1.5)</td>
<td>1 (2.2)</td>
<td>1 (2.2)</td>
</tr>
<tr>
<td></td>
<td>0.3 (0.7)</td>
<td>1 (2.2)</td>
<td>—</td>
</tr>
</tbody>
</table>

$^a$Average: average value for the features included in the categories tumor burden, heterogeneity, and boundaries. Uni (Unidimensional), Bi (Bidimensional). Computation of the rate of decay ($d$) and growth ($g$) of radiomics features using serial radiographs measurement. The eight features discovered in the three signatures (nivolumab, docetaxel, and gefitinib) are generalized and applied to the three cohorts. In the cohort gefitinib, patients had only two evaluations, hence $d$ and $g$ values were not assessable in most patients.
lesions are more likely to progress per iRECIST criteria. Additional details can be found in Supplementary SII.3.

Baseline imaging predictors of radiographic progression in lung lesions treated with gefitinib were surrogates of tumor heterogeneity and were reported previously (39, 45).

Discussion

Using standard-of-care CT images acquired in multicenter clinical trials, machine-learning techniques successfully performed a specific complex task: identifying a pattern of baseline and treatment-induced changes on CT images associated with sensitivity to systemic nivolumab, docetaxel, and gefitinib therapy in patients with a diagnosis of NSCLC (Fig. 4).

Using baseline and first on-treatment assessment (nivolumab: 8 weeks, docetaxel: 8 weeks, gefitinib: 3 weeks), the radiomics signatures output a probability ranging from 0 to 1, corresponding respectively to the highest treatment sensitivity and treatment insensitivity. The innovation of our work compared with the existing literature was that the signatures were dynamic, used a limited subset of features to
reduce overfitting, and signature features were generalized to three treatments evaluated in multicenter studies (46, 47). The eight signature features were reproducible across image reconstruction settings and were robust across tumor sites (Supplementary SII.6). Once designed using machine learning, these signatures can be computed for a given patient using a laptop based on the segmentation of the largest measurable lung lesion by an experienced radiologist on routinely acquired CT scans. Such an analysis could also be incorporated into hybrid imaging modalities such as on the CT portion of a noncontrast PET/CT. Therefore, they could be leveraged—once fine-tuned and optimized in larger cohorts—to guide clinical decisions such as changing systemic therapies at an appropriate time.

These signatures can be understood by both clinical oncologists and radiologists as noninvasive in vivo surrogates of biological changes...
following treatment (Table 1). The eight imaging biomarkers included in the signatures fall into three categories: (i) indicators of change in tumor burden, (ii) indicators of change in tumor spatial heterogeneity, and (iii) characterization of tumor-parenchyma boundaries. Therefore, we can assume that we identified three imaging hallmarks that appear to be prognostic and generalizable quantitative CT radiomics response biomarkers predicting sensitivity to therapies: a decrease in tumor volume, heterogeneity, and tumor-parenchyma invasiveness. The combination of these biomarkers into a signature successfully identified tumors sensitive to cancer treatments. Using serial radiographic measurement and a landmark analysis at 8 months, we demonstrated that a substantial percentage of tumors exhibit an exponential increase in either tumor volume ($g_{\text{volume}}$), tumor spatial heterogeneity of contrast-enhanced images ($g_{\text{GLCM IMC1}}$), or shape irregularity ($g_{\text{shape SI4}}$). We demonstrated that the magnitude of this exponential increase can be leveraged to forecast shorter OS. In sensitive tumor, we observed tumor shrinkage, an increase in tumor homogeneity (decrease in heterogeneity), and a progressive regularization of tumor contours. This is the first demonstration that the evolution of radiomics features deciphering tumor phenotype under systemic therapy selection pressure follows exponential kinetics and coincides with previous work demonstrating that the quantity of tumor increases and decreases exponentially.

Changes in tumor volume predicted treatment sensitivity in all cohorts. However, it played a more important role in the chemotherapies (docetaxel) than in the immunotherapy signature (nivolumab), and was not of importance in the anti-EGFR signature (gefitinib). This is interesting because size-based response criteria were originally developed to assess response of tumors to cytotoxic chemotherapies, with many questioning their general value in assessing targeted molecular agents which are often said to be cytostatic and immunotherapy agents which have been reported to lead to unconventional patterns of response and progression. We demonstrated that temporal changes in intratumoral spatial heterogeneity were associated with sensitivity of NSCLC to anti–PD-1 (nivolumab) as well as anti-EGFR (gefitinib) therapies. Because CT scans are standard of care, noninvasive, informative of the entire tumor burden, and can be performed serially, they are well suited to address spatial and temporal heterogeneity. Tumor heterogeneity is a hallmark of cancer, and some have argued can emerge from the inherent dynamic evolution and adaptation of clones in the presence of drug selection pressure, albeit more likely over prolonged periods of time and not over weeks (48). However, contrast-enhanced CT scans may capture macroscopic patterns of accumulation of iodine linked to tumor neovascularization occurring over shorter time periods. This neovascularization is marked by heterogeneous and excessive blood flow, reduced drug delivery, hypoxia, immune evasion, tumor progression, and metastasis (48). Our modeling framework using the largest measurable lung lesion is supported by studies demonstrating that under drug selection pressure, a similar dynamic is observed in the majority of individual lesions within the same tumor site as well as in lesions with different anatomic locations (49). The incremental value of imaging surrogate-gates of tumor heterogeneity and the kinetics or rates of their evolution should therefore be investigated prospectively as potential candidates to guide precision medicine approaches in systemic therapies with unconventional patterns of response and progression. Pilot results support our findings in the field of immune therapies because lung tumour heterogeneity on contrast-enhanced CT scan has been linked to OS (50) and immune contexture (51).

Early changes in tumor-parenchyma boundaries were strongly associated with tumor sensitivity in patients treated with nivolumab and gefitinib. Although these features are influenced by segmentation (39) and the possibility of change in the shape of lung tumors during respiration, they might capture macroscopically the growth pattern at tumor–lung parenchyma interfaces in NSCLC. Complex lung interfaces are indeed associated with aggressive malignant tumors and poorer survival (52, 53).

In an attempt to generalize the features included in the three signatures, we succeeded in the majority of patients in describing the kinetics of change of the eight radiomics features. We estimated rate constants for both the regression (decrease, $d$) and growth (increase, $g$) of the radiomic features using serial radiographic measurements from multiple timepoints. This is the first demonstration that the effect of a treatment on the rates of growth and/or decay in radiomics features can be estimated. Because OS remains the gold standard for efficacy in oncology, we performed a landmark analysis for OS and could demonstrate that increased $g_{\text{volume}}$ (tumor burden), $g_{\text{GLCM IMC1}}$ (heterogeneity), or $g_{\text{shape SI4}}$ (boundaries invasiveness) were associated with shorter OS. This is the first demonstration that the rate of growth of radiomics features can be used to predict OS (42).

In an ancillary lesion-based analysis conducted on baseline lung lesions, we observed that small invasive lung tumors with high heterogeneity (low RUN—Primitive length uniformity) are more likely to progress per iRECIST criteria and further corroborate the clinical significance of the hallmarks of treatment sensitivity used in the nivolumab signature.

Our study has limitations. This is a proof-of-concept study with relatively small sample size. The sample was representative of a population of NSCLC in a large multicenter clinical trial. Using a multicenter clinical trial and randomization of patients in training and validation sets reduced the risk of overfitting. There was no apparent selection bias with the covariates balanced between included and excluded patients, and between training and validation cohorts for each treatment. However, there was a selection based on the presence of a measurable lung lesion. The high frequency of measurable lung lesions in NSCLC in this series made our model applicable to 61% of patients with refractory NSCLC treated with nivolumab and 100% of patients with early stage NSCLC treated with gefitinib. The overall classification performance of the nivolumab signature was better than that of the gefitinib signature. However, the dataset was unbalanced with a minority (16%) of tumor sensitive to treatment. Hence, the final machine-learning model included one single feature to accurately predict treatment sensitivity because more complex models would have been overfitted. The radiomics pipeline is complex, which makes the selection and identification of imaging biomarkers difficult to be widely adopted in nonacademic institutions. Nonetheless, we extensively described our methodology and clinically relevant features. In addition, we have identified a limited subset of imaging biomarkers that could be easily implemented and computed on any laptop. The study was not designed to evaluate how various time intervals alter feature selection and classification. However, delta features identified at 3 weeks (NCT00588445) were generalizable at 8 weeks (NCT01642004, NCT01721759).

The radiomics signatures were applicable in selected patients with measurable lung tumors reaching predefined clinical and imaging quality criteria at baseline and at first CT evaluation (Fig. 1). The output to be predicted by radiomics signatures was PFS rather than OS because OS suffered from limitations such as unbalanced number of candidates.
events (nivolumab: 33 deaths occurring with a median follow-up of 8.6 months), the occurrence of crossovers with new treatment lines in patients experiencing disease progression (nivolumab and docetaxel), and the excellent outcome of resectable early-stage NSCLC (gefitinib). Perhaps more or different information would be obtained with evaluations earlier than 8 weeks (nivolumab and docetaxel), as demonstrated by the value of a 3-week evaluation (gefitinib).

In conclusion, this study is a proof of concept that AI support could provide clinicians an early indication of the likelihood of success of treatment with the new generation of systemic anticancer therapies using conventional imaging techniques. Computers excelled in mining and integrating large amounts of data from a quantitative CT analysis of a single lung lesion segmented by an expert radiologist. Using this data, early changes in tumor imaging phenotype on standard-of-care CT scan were translated into a quantitative and synthetic signature to predict treatment sensitivity. Treatment sensitivity was associated with changes in the interface between lung tumor and normal lung parenchyma, as well as heterogeneity of the lung tumor. Once further prospectively validated, these signatures could be used clinically to enhance the strategic decision-making of a practicing clinical oncologist optimizing precision treatment. Consequently, AI-generated on-treatment signatures could allow for more accurate treatment decision-making which could constitute the basis for the implementation of adapted treatment guided by quantitative CT scan interpretation in patients with NSCLC treated with systemic therapies.

Disclosure of Potential Conflicts of Interest

M. Fronheiser is an employee for Bristol-Myers Squibb. S. Du is an employee for Bristol-Myers Squibb. W. Hayes is an employee for and holds ownership interest (including patents) in Bristol-Myers Squibb. D.K. Leung is an employee for Bristol-Myers Squibb. A. Roy is an employee for and holds ownership interest (including patents) in Bristol-Myers Squibb. L.H. Schwartz is a paid advisory board member for Roche and Novartis, and reports receiving commercial research grants from Merck and Boehringer Ingelheim. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L.H. Schwartz, B. Zhao


Writing, review, and/or revision of the manuscript: L. Dercle, M. Fronheiser, L. Lu, S. Du, W. Hayes, D.K. Leung, A. Roy, A.T. Fojo, L.H. Schwartz, B. Zhao

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D.K. Leung, P. Guo, L.H. Schwartz, B. Zhao


Acknowledgments

Authors acknowledge financial support from the NIH (U01 CA225431) and Bristol-Myers Squibb. L. Dercle’s work was partially funded by grants from Foundation Philanthropia and Fondation Nuovo-Soldati. The content is solely the responsibility of the authors and does not necessarily represent the funding sources.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 7, 2019; revised November 27, 2019; accepted January 22, 2020, published first March 20, 2020.

References


Clinical Cancer Research

Identification of Non–Small Cell Lung Cancer Sensitive to Systemic Cancer Therapies Using Radiomics

Laurent Dercle, Matthew Fronheiser, Lin Lu, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-19-2942

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2020/07/07/1078-0432.CCR-19-2942.DC1

Cited articles
This article cites 50 articles, 4 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/26/9/2151.full#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/26/9/2151.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/26/9/2151.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.